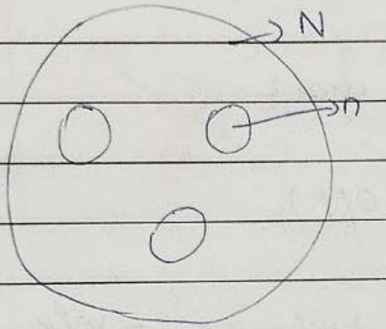
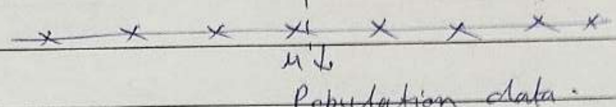
Q. Why sample variance is divided by $n-1$.

$$\mu = \sum_{i=1}^{N} \frac{x_i}{N} \qquad\qquad \bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$$\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N} \qquad\qquad \boxed{s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}$$

Unbiased Estimation





→ When we calculate μ & $\bar{x}$ in the above data it would be similar $\mu \approx \bar{x}$

→ In case of skewed data we would be picking up data points and the $\bar{x}$ would be too small compared to μ.

→ Inorder to address this issue we use $n-1$ in case of sample variance.

→ This whole terminology is also called as 'Unbiased Estimation'.

DATE / /
PAGE

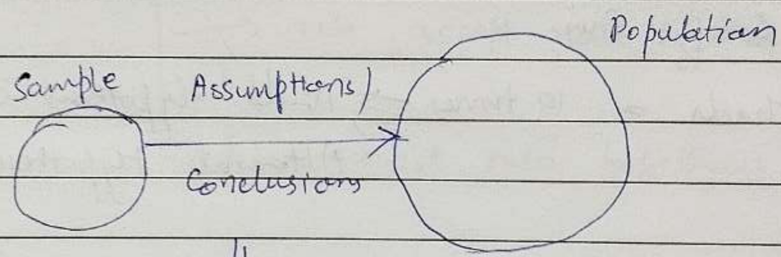
DAY - 5 STATS

Inferential Statistics

1. Hypothesis testing
2. p - value
3. Confidence Interval
4. Significance value

1. Hypothesis Testing

Inferential Stats



⇓

For validating these assumptions we use hypothesis testing.

i) Null Hypothesis :- Null hypothesis will always take the default value. Eg:- If a person has committed crime by default null hypothesis is person is not a criminal.

Experiment
[Coin is fair or not] → Null Hypothesis - Coin is fair
$P(H) = 0.5$ $P(T) = 0.5$ ↑ Default value

ii) Alternate Hypothesis - (Always) opposite, Eg:- Coin is not fair, Person has committed the crime

③ Perform Experiments :-

Tossing coins $\boxed{100 \text{ times}}$

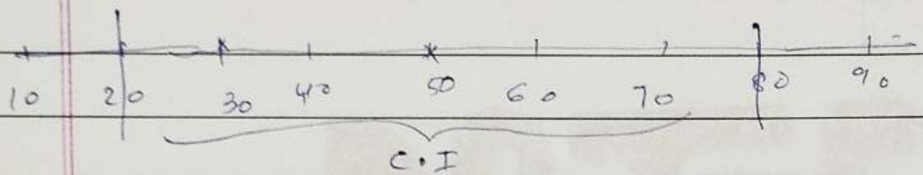
⇓

50 times head → Fair

60 times head ⇒ Fair

70 times head ⇒ Fair

80 times head ⇒ Not fair



* We will ask the Domain experts what will be the $\boxed{\text{Confidence Interval}}$ for coin to be fair.

$$CI = [20 - 80]$$

⇓

Coin is fair

⇒ For Heads = 10 times →, Null Hypothesis is rejected
Alternate Hypothesis is accepted

⇒ If number is within CI we fail to reject Null Hypothesis.

→ We reject null hypothesis when [outside C.I]

Conclusions

EXAMPLE 2 :

(i) Null Hypothesis – Not a criminal (Default value)

(ii) Alternate Hypothesis :- Person is a criminal

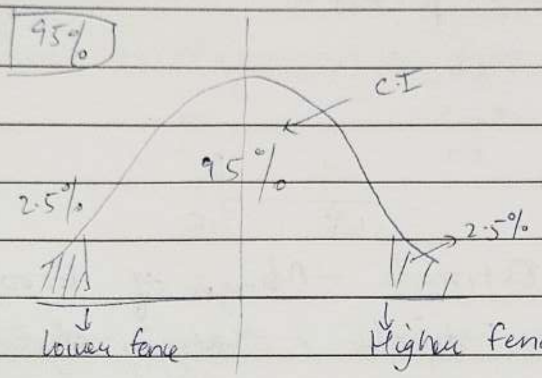
(iii) Experiment / Proof :- DNA, finger print, weapons, eye witness, footage

⇓

Judge ⇒ conclusion

Defined by domain expert.

Confidence Interval (CI) Significance value

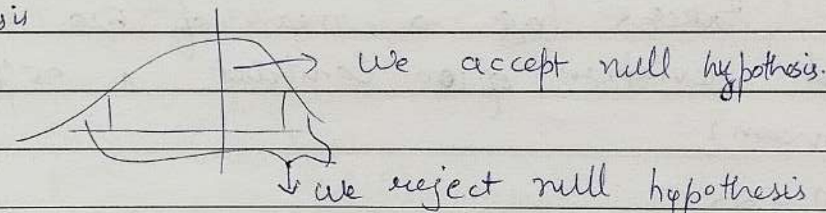


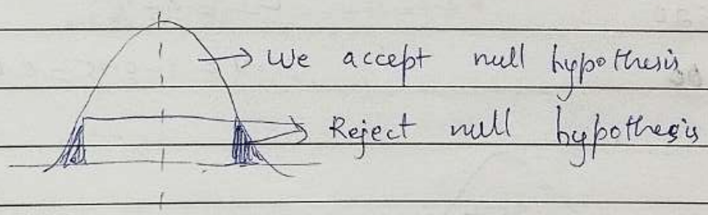
$CI = 95\%$

$S.V = 1 - 0.95 = 0.05$

Eg:- In case of vaccine testing CI should be extremely low to use it. Because lives matter

- When value b/w [lower fence - Higher fence] we accept null hypothesis





* Point Estimate :- The value of any statistic that will estimates the value of a parameter is called point estimate.

$\bar{x} \rightarrow \mu$ → Using $\bar{x}$ we can make conclusions on μ in case of inferential stats.

$\bar{x}$ ↓ Statistics

μ ↓ Parameter

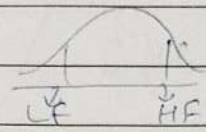
Using $\bar{x}$ statistic estimating value of parameter μ.

→ In the ~~exp~~ example above $\bar{x}$ is the point estimate.

Point Estimate ± Margin of Error = Parameter ⇒ population Mean

$$\boxed{\text{Point Estimate} \pm \text{Margin of Error} = \text{Parameter} \Rightarrow \text{population mean}}$$

With respect to CI,



Lower Fence :- Point Estimate − Margin of Error
Higher Fence :- Point Estimate + Margin of Error

$$\boxed{\text{Margin of Error}} = \boxed{Z_{\alpha/2} \times \boxed{\frac{\sigma}{\sqrt{n}}}} \rightarrow \text{Standard Error}$$

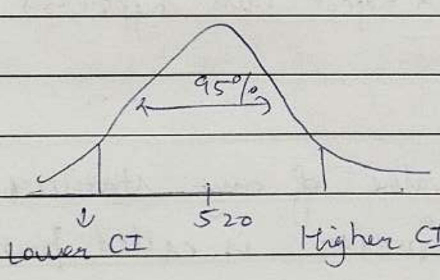
$\alpha \rightarrow$ Significance value

Q. * On the quant test of CAT exams a sample of 25 test takers has a mean of 520 with a ~~sample~~ population standard deviation of 100. Construct a 95% CI about the mean?

Ans. $n = 25$ $\qquad$ $CI = 95\%$

$\bar{x} = 520$ $\qquad$ $SV = 1 - 0.95 = 0.05$

$\sigma = 100$



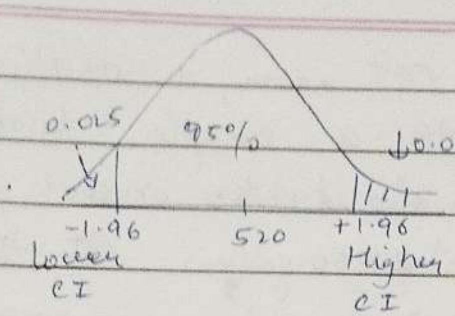
Lower CI = Point estimate − Margin of Error

$$= 520 - Z_{0.05/2} \boxed{\frac{100}{\sqrt{25}}}$$

$$= 520 - Z_{0.025} \times 20$$

$$= 520 - 1.96 \times 20$$

$$= \underline{480.8}$$



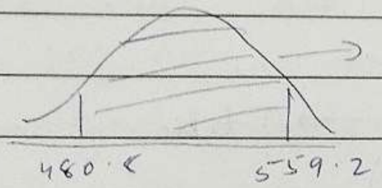
$\boxed{1 - 0.025 = 0.975}$

↓

Standard deviation in Z table is 1.96.

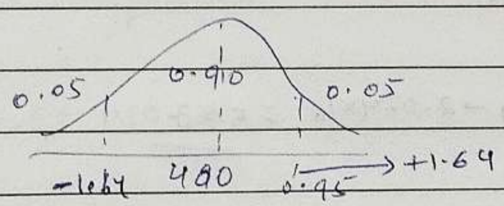
for 0.025 $\sigma = -1.96$

~~Lower~~ Higher fence = $520 + 1.96 \times 20 = 559.2$



→ Failed to reject Null Hypothesis

Q. $\bar{x} = 480$, $\sigma = 85$, $n = 25$, $CI = 90\%$ (0.90)

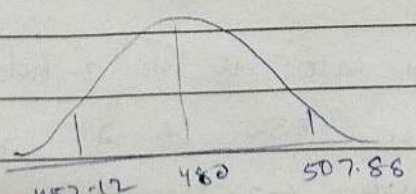


$$\begin{array}{r} 164 \\ \times\ 17 \\ \hline 1148 \\ 164 \\ \hline 27.88 \end{array}$$

Lower CI = $480 - Z_{0.05} \times \frac{85}{5}$

$= 480 - 1.64 \times 17 = 480 - 27.88 = 452.12$

Higher CI = $480 + 27.88 = 507.88$



Q. On the quant test of CAT exam, a sample of 25 test takers has a mean of 520, with a sample standard deviation of 80. Construct 95% CI about the mean.

Ans. $\bar{x} = 520$, $n=25$, $s=80$, $CI = 95\%$, $SV = 0.05$

Since σ is not given,

$$\boxed{\text{Point estimate} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)}$$

t - test

Degree of freedom $= n-1 = 25-1 = 24$ → T-table

lower CI $= 520 - t_{\frac{0.05}{2}} \times \left(\frac{80}{5}\right)$

$= 520 - 2.064 \times 16$

$= 486.976$

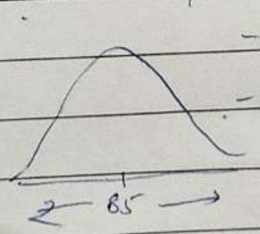
Higher CI $= 520 - 2.064 \times 16 = 553.024$

(1) 1 Tail and 2 Tail Test

Q. Colleges in Town A has 85% placement rate. A new college was recently opened and it was found that a sample of 150 students had a placement rate of 88% with a standard deviation of 4%. Does this college has a diff placement rate with 95% C.I ?

Ans.

- in this case it is 2-tailed test.



- If it was 85% it is 2-tailed test
- If it was <85% it is 1-tailed test.

Hypothesis Testing

Q. A factory has a machine that fills 80 ml of baby medicines in a bottle. An employee believes that the average amount of baby medicine is not 80ml. Using 40 samples he measures the average amount by the machine is to be 78 ml with standard deviation of 2.5. ~~State null and alternate hypothesis.~~

a) State null and alternate hypothesis

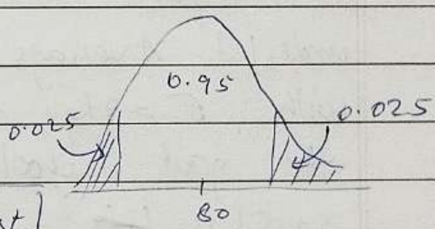
b) At 95% CI, is there enough evidence to support machine is working properly or not.

Ans: a) $n = 40$, $\bar{x} = 78$, $s = 2.5$

Null hypothesis − Mean = 80, $\mu = 80$ml

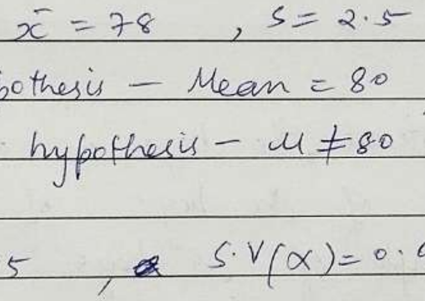
Alternate hypothesis − $\mu \neq 80$

b) CI = 0.95, S.V(α) = 0.05



Note: when $n \geq 30$ or given $\sigma \rightarrow$ Z test
$n < 30$ & give $s \rightarrow$ t test

Here we have to use Z-test,



These, $\sigma = -1.96$ & 1.96 are $\boxed{\text{decision boundary}}$

Calculate test statistics (z-test)

$$Z = \frac{\bar{x} - \mu}{\boxed{s/\sqrt{n}}} \rightarrow \text{Standar Error}$$

$$= \frac{78 - 80}{2.5} = \frac{-2}{\frac{2.5}{\sqrt{40}}} = \frac{-4}{5} = -0.8 \times \sqrt{40}$$

$$= -5.05$$

· Conclusions

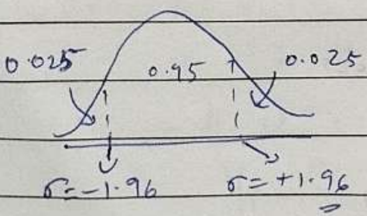
Decision Rule :- If $z = -5.05$ is less than -1.96 or greater than $+1.96$ we reject null hypothesis; reject null hypothesis with 95% CI.

Reject the Null hypothesis $\{$ Theres is some fault in the machine

Q. A complain was registered, the boys in a govt. school are underfed. Average wt. of the boys of age 10 is 32kgs. with $\sigma = 9$ kg. A sample of 25 boys were selected from the govt. school and the avg wt was found to be 29.5 kgs.
with CI - 95%, check if it is true or false.

Ans. Z-test will be used here

$\mu = 32$, $n = 25$
$\sigma = 9$ $\quad \bar{x} = 29.5$ kg.



$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{(29.5 - 32)5}{9}$$

$$= \frac{-2.5 \times 5}{9}$$

$$= \frac{-12.5}{9} = -1.39$$

Step 1

$H_0 : \mu = 32$
$H_1 : \mu \neq 32$

~~Cond.~~

→ $-1.39 > 1.96$, ∴ we ~~accept~~ fail to reject the null hypothesis.
Boys are being fed properly

Q. A factory manufactures cars with a warranty of 5 years o more on the engine and transmission. An engineer believe that engine or transmission will malfunction in less than 5 years. He tests a sample of 40 cars and finds the a time to be 4.8 years with a standard deviation of 0

1) State the null & alternate hypothesis.
2) At a 2% significant level, is there enough evidence t support the idea that the warranty should be revised

Ans: Step 1

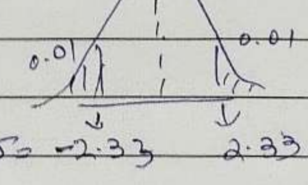
$H_0 : \mu \geq 5$

$H_1 : \mu < 5$

2) $n = 40$, $\bar{x} = 4.8$, $s = 0.5$

S.V $\alpha = 0.02$, $CI = 0.98$

0.99



~~.02~~

Z-test

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{4.8 - 5}{\frac{0.5}{\sqrt{40}}}$$

$$= \frac{-0.2}{0.5} \times \sqrt{40} = -0.4 \times \sqrt{40}$$

$$z = -2.529$$

→ In this case we will reject null Hypothesis.

Q. In the population the average IQ is 100 with a standard deviation of 15. A team of scientists wants to test a new medication to see if it has a +ve or -ve effect, or no effect at all. A sample of 30 participants who have taken the medication has a mean of 140. Did the medication affect intelligence {95% CI}

Ans: $\mu = 100$, $\sigma = 15$

$n = 30$, $\bar{x} = 140$

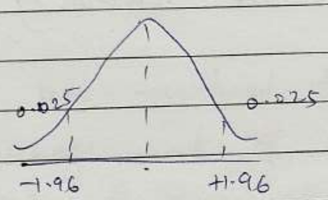
z test.

$H_0: \mu = 100$

$H_1: \mu \leq 100$, $\mu = 140$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{40}{\frac{15}{\sqrt{30}}}$$

$$= \frac{8 \times \sqrt{30}}{3} = 14.6059$$



→ We have rejected null hypothesis here since z is more than +1.96.